

## Bachelor-/Masterarbeit

### Erarbeitung von Verfahren zur Semantischen Anreicherung technischer Dokumente auf Basis von Ontologien aus strukturierten technischen Daten

#### Hintergrund

Die Heidelberger Druckmaschinen AG (im folgenden „HDM AG“) hat ein Serviceinformationsportal für technische Dokumente und Daten seit 2009. Es wird weltweit von mehr als 2.000 Technikern genutzt. Die Dokumente dafür werden zum größten Teil von technischen Redakteuren mit professionellen Werkzeugen erstellt und mit ihren Übersetzungen in Englisch, Französisch und Spanisch sowie Metadaten auf dem Serviceinformationsportal bereitgestellt. Im Sommer 2013 hat die HDM AG die Fa. Gallus in der Schweiz, die Flexodruckmaschinen herstellt und vertreibt, zu 100% übernommen. Deren technische Dokumente sind bisher nicht im Serviceinformationsportal verfügbar. Längerfristiges Ziel ist es, mit minimalem manuellem Aufwand die technischen Dokumente von Gallus in der gleichen Systematik in das Serviceinformationsportal zu bekommen.

#### Ziel der Arbeit

Ziel ist es, nachzuweisen, ob man mit den Methoden der semantischen Anreicherung ausreichend und vollständige Metadaten für technische Dokumente generieren kann, um alle technischen Begriffe bzw. Objekte als solche zu erkennen. Dazu sollen Prozesse und automatische Verfahren entwickelt werden, wie die mit Hilfe einer Software zur semantischen Anreicherung („Luxid“ der Fa. Temis Heidelberg) und strukturierten technischen Daten aus dem SAP-System der HDM AG notwendige Ontologien gebildet und Prozesse entwickelt werden, die zur automatischen Generierung von Metadaten für die technischen Dokumente dienen.

#### Ablauf

In einer ersten Stufe soll auf Basis strukturierter technischer Daten aus dem SAP-Verwaltungssystem der HDM AG Ontologien mit Hilfe der Software „Luxid“ für semantischen Anreicherung genutzt bzw. weiterentwickelt werden. Dies umfasst Produktbezeichnungen, Artikelnummern und –bezeichnungen, Fehlermeldungen etc. Diese Ontologien sollen in einem weiteren Schritt für die Erkennung der Begriffe in den vorhandenen technischen Dokumenten einer typischen Baureihe der HDM AG genutzt. Ziel muss es sein, so viele technische Begriffe wie möglich zu erkennen. Außerdem soll mit den genannten Werkzeugen ein Prozess entwickelt werden, der bedeutende „Wissensgebiete“ in den Dokumenten und in den Stücklisten identifiziert, die auf funktionalen Baugruppen oder Anwendungsbereichen der Maschine beruhen. Diese Wissensgebiete sollen später als Suchfilter für technische Informationen dienen.

#### Voraussetzungen

Sie sollten gute Programmierkenntnisse und Kenntnisse in Datenmodellen (Excel/Access) haben, um die Anpassungen bei den Ontologien machen zu können. Darüber hinaus erwarten wir gute Kenntnisse in den Grundlagen der semantischen Anreicherung.

#### Rahmenbedingungen

Die Dauer der Arbeit beträgt mindestens 3 Monate ohne die schriftliche Ausarbeitung. Die monatliche Vergütung beträgt 500 - 700 € abhängig ob Bachelor- oder Masterarbeit. Die HDM AG kann nur Studenten mit einer in der Studienordnung vorgeschriebenen Bachelor- oder Masterarbeit unterstützen. Die Betreuung der Arbeit und die Bereitstellung der SAP-Daten erfolgt durch einen Mitarbeiter der HDM AG und des AIFB.

#### Sie wollen sich bewerben?

Dann schicken Sie uns bis 01.03.2015 eine E-Mail mit Ihrer Motivation und aktuellem Notenauszug an [patrick.philipp@kit.edu](mailto:patrick.philipp@kit.edu)